DEPARTMENT OF STATISTICS

The Ohio State University

OSU

COLUMBUS, OHIO

LEVEL II

(12)

6 OPTIMIZING METHODS IN SIMULATION

by

10 J. S. Rustagi

DTIC
ELECTE
AUG 14 1981
S D
E

14 TR-239

Technical Report No. 239
Department of Statistics
The Ohio State University
Columbus, Ohio 43210
July 1981

(12)

Supported in part by Contract No. N000-14-78-C-0543 (NR 042-403)
by Office of Naval Research. Reproduction in whole or in part
is permitted for any purpose of the United States Government.

426221

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 9	2. GOVT ACCESSION NO. AD-A102846	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Optimizing Methods in Simulation		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER 239
7. AUTHOR(s) J. S. Rustagi		8. CONTRACT OR GRANT NUMBER(s) N000-14-78-C-0543
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics The Ohio State University 1958 Neil Avenue, Columbus, OH 43210		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-403
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of Navy Arlington, Virginia 22207		12. REPORT DATE August, 1981
		13. NUMBER OF PAGES 41
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Simulation, optimizing techniques, response surface, stochastic approximation, optimal design.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Simulation techniques utilize methods of optimization in several aspects. Validation of simulation models, estimation of models and design of simulation experiments require optimization. Methods of optimization are discussed with special applications to simulation. Criteria of optimization, classical methods of optimization, numerical methods, and optimal search procedures are discussed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

1. Introduction

Simulation of various physical, biological or social complex systems allows us to develop elaborate models for them and helps in the process of making valid inferences from them. There are many situations in which systems can not be easily described in a compact form for analysis and prediction. The modern computer simulations allow us to represent such systems by series of simpler models and thus help us in providing reasonable solutions to complex problems.

A schematic representation of the simulation strategy for developing models of complex systems has been given by Ziegler et. al. (1979).

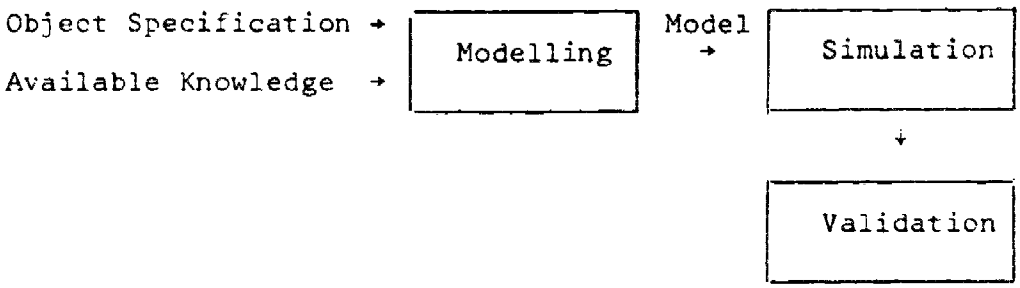


Figure 1

The validation of models requires some sort of optimization. One has to provide criteria of optimization and possible techniques to achieve that optimization to

complete the process of validation.

Optimizing techniques are also required in other aspects of simulation experiment such as in their design and ultimate analysis. The method of optimization form a vast body of knowledge spreading over several fields. We have classified optimizing methods broadly in the following categories and have arranged the list of references in that order.

A. Classical optimizing techniques;

B. Numerical procedures;

C. Mathematical programming methods;

D. Stochastic approximation methods;

E. Optimum seeking methods and response surface methodology;

F. Optimal Design Theory;

G. Miscellaneous methods.

In this brief account, we emphasize those optimization techniques which are of potential use in simulation methodology. We shall concentrate here on Optimizing Criteria, Classical Methods, Numerical Methods, Optimal Search Procedures, Response Surface Methods and Optimal Designs of Regression Experiments.

However, technical references are provided on various other optimization techniques for the interested reader.

2. Optimizing Criteria

Optimization is basically dependent on the criteria used in a given situation. The same problem may lead to different solutions depending upon the criteria of optimality utilized. The criteria depend on the nature of the problem and are many times dictated by practical considerations. Consider the case of least square estimation of parameters in hypothesized models. The criterion of minimizing the sum of squares of residuals, was dictated more as a mathematical convenience than from heuristic point of view. It allows simple mathematical solution is most cases. However, if the criterion of optimality is chosen to be that of minimizing the sum of absolute deviations of residuals, the mathematical simplification is minimal and recourse has to be made to numerical solutions. It may be highly important to select the "right" criterion of optimality in a given situation.

There does not seem to be a simple and logical approach of choosing among a class of competing criteria of optimality for a given problem. Experience and intuition in a given setting may be the ultimate judge for proper selection. In many situations, however, more is known about the comparative properties of the optimality

criteria and the experimenter is guided by such considerations to select the appropriate criterion. We shall discuss some of the most commonly used criteria in this section.

Least Squares Criterion

One of the most common criterion used in validation of models is that of least squares. Given the realization of the process from simulations or actual observations, the observed and the expected value under the assumed models are compared. If the sums of squared deviation is minimized, this method provides the unknown parameters of the model. Various other criteria such as sum of absolute deviations or weighted least squares criterion are also in use. The criterion to be chosen heavily depends on the experimental situation.

Example (Milstein (1979))

In a biochemical process, the equations of the process are described by the following

$$\frac{d\underset{\sim}{x}}{dt} = f(\underset{\sim}{x}, \underset{\sim}{k}),$$

$$\underset{\sim}{x}(0) = \underset{\sim}{c}_i,$$

$$i = 1, 2, \ldots, \ell,$$

and the vector $\underset{\sim}{x}$ is n-dimensional with nonnegative components, $\underset{\sim}{k}$ is a vector of parameters having p unknown components, f is a vector function. The vector $\underset{\sim}{c}$ represents the given initial condition. Let the data be given by $y^s(t_r)$ at rth time point t_r and let the corresponding value of x be given by $x(\underset{\sim}{k}, t_r)$. Let $\underset{\sim}{W}_r$ be the matrix of known weights, then a common measure of the discrepency between the data points y and the trajectories can be the following

$$F(\underset{\sim}{k}) = \sum_{s=1}^{\ell} \sum_{r=1}^{M} [\underset{\sim}{y}^s(t_r) - \underset{\sim}{x}^s(\underset{\sim}{k}, t_r)]' \underset{\sim}{W}_r [y^s(x_r) - \underset{\sim}{x}^s(\underset{\sim}{k}, t_r)$$

M is the number of points chosen. The object will be to determine the unknown parameters $\underset{\sim}{k}$ which can be obtained by using the criterion of minimizing $F(\underset{\sim}{k})$.

A computer algorithm is given in terms of an iterated numerical procedure starting with a first guessed value of $\underset{\sim}{k}$ by Milstein (1979).

In the context of design of experiments, which are highly pertinent to the simulation experiments, we discuss a few criteria which are in commonly use.

Consider the model,

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

where $\underset{\sim}{y}$ is the observation vector in n-dimensions, $\underset{\sim}{X}$ is a $n \times p$ design matrix, $\underset{\sim}{\beta}$, a $p \times 1$ vector of unknown parameters and $\underset{\sim}{\varepsilon}$, a $n \times 1$ vector of residuals. If we use least squares method to estimate $\underset{\sim}{\beta}$, it is well known that we optimize $\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} = (\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{\beta})$ leading to the optimal estimates of $\underset{\sim}{\beta}$ as given by

$$\hat{\underset{\sim}{\beta}} = (X'X)X'^{-1}\underset{\sim}{y}$$

In the problem of finding optimum $\underset{\sim}{X}$ such that the parameter $\underset{\sim}{\beta}$ is estimated optimally, one considers the covariance matrix of $\hat{\underset{\sim}{\beta}}$ given by

$$V(\hat{\underset{\sim}{\beta}}) = (X'X)^{-1}\sigma^2$$

where $\underset{\sim}{\varepsilon}$ is assumed to have means zero and covariance $\sigma^2\underset{\sim}{I}$.

By an experimental design, we mean the choice of levels of X. Consider the case in one dimension for present and assume that there are n observations available. We are interested in knowing the method of allocation of these observations to the various levels of x's. That is, the problem is find levels $x_1, x_2, \ldots, x_k$ to be repeated $n_1, n_2, \ldots, n_k$ times such that $n_1 + n_2 + \ldots + n_k = n$. The set of x_i's with n_i's is called the <u>design of an experiment</u>. In place of integers n_i, we can use fractions

$$p_1, p_2, \ldots, p_k$$

with $\frac{n_i}{n} = p_i$ and $\sum p_i = 1$. The collection of x_i's with p_i's describes generally a discrete probability measure. The theory of optimal design of experiments is concerned with obtaining such a measure so as to optimize some objective function of the parameters in the assumed model for the experiment.

There are several optimality criteria in the case of regression design of experiments and they are given in terms of the matrix $\underset{\sim}{X}'\underset{\sim}{X}$. Suppose $\underset{\sim}{X} = (\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n)$, with $\underset{\sim}{x}_i$, $i = 1, 2, \ldots, n$ being p-vectors and let $x_i \in X$.

Criterion of G-Optimality

It is also known as the criterion of minimax optimality.

Find x_i such that

$$\min_{\substack{\underset{\sim}{x}_i \\ i=1,2,\ldots,n}} \; \max_{\underset{\sim}{x} \in} \{X'(X'X)^{-1}X\}$$

Criterion of D-Optimality

In this criterion, we find x_i such that determinant of the matrix $X'X$ is maximized. That is, find x_i, such that we have

$$\max_{x_i} \det (X'X)$$

Criterion of A-Optimality

Find $\underset{\sim}{x}_i$ such that

$$\min_{x_i} \operatorname{trace} (X'X)^{-1}$$

Criterion of E-Optimality

This criterion is concerned with finding x_i such that minimum eigenvalue of X'X is maximized. That is,

$$\max_{x_i}(\text{min eigenvalue of } \underset{\sim}{X}'\underset{\sim}{X})$$

Many other kinds of optimality criterion in the context of design of regression experiments have been discussed in the literature, for reference, see Federov (1972).

Integrated Mean Square Error Criterion

Rcently Brown (1979) has proposed the integrated mean square error as an optimization criterion in the context of linear inverse.

This criterion has been used in other contexts as well, see Tapia and Thompson (1978). A common measure of discrepency between the observed and expected value is obtained in terms of mean squared errors (MSE).

Consider the model,

$$E(Y|x) = \alpha + \beta x$$

and

$$V(Y|x) = \sigma^2$$

Let $L \leq x \leq U$, be the interval of possible x values. The MSE(x) is the mean squared error of x as obtained from y. Let w(x) be a weight function. Then Integrated Mean Square Error is defined as

$$\text{IMSE} = \int_L^U \text{MSE}(x)\, W(x)dx$$

In calibration problems, Brown has shown that optimization of IMSE gives much better results as compared to simply minimizing MSE. In case, no special form of the weight function W(x) is suggestible from the problem, W(x) may be taken to be uniform over the range (L, U).

3. Classical Methods of Optimization

The basic problem of optimization is concerned with finding a value x_0 in a finite dimensional set A, for which a function f(x) defined on the set A, attains a maximum or a minimum. If A is a finite set, the minimizing and maximizing values always exist. They need not exist when A is not finite.

$$\text{Suppose } f(x) = \begin{cases} 1, & x = 0, \\ x, & x > 0. \end{cases}$$

Then the function $f(x)$ defined our $x \geq 0$, the non-negative part of the real line does not have a minimum which can be attained. The ideas of infimum and supremum are introduced to take care of such a situation.

Define supremum of $f(x)$ or sup $f(x)$ by the least value of λ such that

$$f(x) \leq \lambda. \quad \text{for all } x \in A.$$

Similarly infimum of $f(x)$ or inf $f(x)$ is defined by the largest value λ such that $f(x) \geq \lambda$.

An important result in this regard is given by the following theorem.

Theorem 3.1. If $f(x)$ is continuous and the set A is finite and closed interval then $f(x)$ attain its minimum or maximum (extrema) values in A.

For proof, see any book on calculus, for example, Whittle (1971).

The necessary and sufficient conditions for extrema are given by the following theorems, usually available in standard calculus books.

Theorem 3.2. (Necessary Conditions for an extremum) If the derivative $\frac{\partial f}{\partial x}$ exist at an interior point, x_0, of the set A, and if x_0 is an extremum point, then $\frac{\partial f}{\partial x} = 0$ at $x = x_0$.

Define the Hessian of a function f(x) by the matrix of second order partial derivations as follows.

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \, x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Theorem 3.3. (Sufficient Condition for an extremum)

The sufficient condition that f(x) has a maximum (minimum) at an interior point $x_0 \in A$ is that H exist and be negative definite (positive definite).

The proofs require expanding the function f(x) with the help of Taylor's theorem using H. For details see, Whittle (1971).

Constrained Optimization

In finding extrema of a function f(x) over the set A, these may be additional constraints added such as by the condition, g(x) = b. Essentially the constraints introduce a subset of the set A over which f(x) should be optimized. The case when the constraints are introduced by inequalities is dealt with by mathematical programming methods.

The method of Lagrange multipliers has been used extensively for solving constrained optimization problems. The method requires optimizing

$$f(x) + \lambda g(x)$$

where λ is some unknown constant. If the number of constraint equation is more than one, Lagrange's method requires optimizing

$$f(x) + \underset{\sim}{\lambda}'\underset{\sim}{g}(x)$$

where $\underset{\sim}{g}$ is the vector of function given and the vector $\underset{\sim}{\lambda}$ is unknown. For an extensive discussion, see Whittle (1971).

4. Numerical Methods of Optimization

By the very nature of the simulation process, numerical methods are necessary for optimizing techniques for simulation models. In the case of functions of one variable, it may sometimes be easy to graph the function and then obtain the optimizing value. In the case of several variables, the process involves large numbers of calculation and may exceed the limit of computers.

The optimization of functions in many cases reduces to finding the solutions of equations since the extremizing values are given by the derivatives or partial derivatives if they exist. We first consider methods of solving an equation of the type,

$$f(x) = 0 \tag{4.1}$$

General methods for solutions are available in textbooks of numerical analysis, for example see Ralston (1965). We first define Lagrange polynomials which are used in interpolation. Lagrange polynomial of (n-1)-th degree are defined by

$$\ell_j(x) = \frac{p_n(x)}{(x-a_j)p_n'(a_j)}, \qquad j = 1, 2, \ldots, n \tag{4.2}$$

where

$$p_n(x) = (x-a_1)(x-a_2) \ldots (x-a_n) \tag{4.3}$$

is a polynomial of nth degree with given contants a_1, a_2, $\ldots$, a_n.

$p_n'(a_j)$ gives the derivative of the polynomial $p_n(x)$ at a_j. For example, Lagrange polynomials of order 3 are given by

$$\ell_1(x) = \frac{(x-a_2)(x-a_3)}{(a_1-a_2)(a_1-a_3)} \tag{4.4}$$

$$\ell_2(x) = \frac{(x-a_1)(x-a_2)}{(a_2-a_1)(a_2-a_3)} \tag{4.5}$$

$$\ell_3(x) = \frac{(x-a_1)(x-a_2)}{(a_3-a_1)(a_3-a_2)} \tag{4.6}$$

<u>Iterative procedure for roots of the equation</u>, $f(x) = 0$.

Suppose inverse of the function f exists. Let $y = f(x)$ so that $x = f^{-1}(y) = g(y)$. We are looking for $g(0)$ which will be the root α. That is, $g(0) = \alpha$.

The Lagrange interpolation formula gives an approximation for $g(y)$ by $h(y)$, denoted by, $g(y) \approx h(y)$.

$$\begin{aligned} h(y) &= \sum_{j=1}^{n} \ell_j(y) g(y_j) \\ &= \sum_{j=1}^{n} \ell_j(y) x_{i-j+1} \end{aligned} \tag{4.7}$$

where $g(y_j) = x_{i-j+1}$, given the points, y_1, y_2, $\ldots$, y. An approximation of α by x_{i+1} is given by $h(0)$. That is,

$$x_{i+1} = \sum_{j=1}^{n} \ell_j(0) \quad x_{i+1-j} \tag{4.8}$$

Notice that

$$\ell_j(0) = \frac{(-1)^n y_1 y_2 \cdots y_{j-1} y_{j+1} \cdots y_n}{(y_j-y_1)(y_j-y_2)\cdots(y_j-y_{j-1})(y_j-y_{j+1})\cdots(y_j-y_n)} \tag{4.9}$$

The equation (4.9) gives an n-point iteration process. That is given x_i, x_{i-1}, ..., $x_{i-(n-1)}$, we can find x_{i+1}. Or the n-point iteration function is given by

$$x_{i+1} = F_i(x_i, x_{i-1}, \ldots, x_{i-(n-1)}) \tag{4.10}$$

Most iteration procedures use only one point iteration and the same function for iteration. That is,

$$x_{i+1} = F(x_i) \tag{4.11}$$

There are many methods of iteration. We shall discuss here the most commonly used methods such as those of Newton-Raphson.

Newton-Raphson Procedure

In this procedure , we use

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{4.12}$$

Using the approximation of $f'(x_i)$ by

$$f'(x_i) \approx \frac{f(x_i)-f(x_{i-1})}{x_i-x_{i-1}} \tag{4.13}$$

we have $x_{i+1} = x_i - \dfrac{f(x_i)(x_i-x_{i-1})}{f(x_i)-f(x_{i-1})} =$

$$= \frac{x_i f(x_{i-1})}{f(x)-f(x_{i-1})} - \frac{x_{i-1} f(x_i)}{f(x_{i-1})-f(x_i)} \tag{4.14}$$

The above two-point iteration is known as the <u>secant method</u>.

In the case of several equations, a generalized one-point Newton-Raphson iteration procedure can be similarly described. Let $x = \binom{x_1}{x_2}$ be a two dimensional vector. Let

$$f_1(\underset{\sim}{x}) = 0 \qquad \text{and} \qquad f_2(\underset{\sim}{x}) = 0$$

be the two simultaneous equations to be solved. Then the Netwon-Raphson iteration requires the following:

$$\underset{\sim}{x}_{i+1} = \underset{\sim}{x}_i - \left. \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \\ \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}^{-1} \begin{pmatrix} f_1(\underset{\sim}{x}_1) \\ \\ f_2(\underset{\sim}{x}_2) \end{pmatrix} \right|_{\underset{\sim}{x}=\underset{\sim}{x}_i} \tag{4.15}$$

<u>Gradient Method</u>

Gradient method was introduced by Cauchy in 1847. This method utilizes the gradient of the function $f(\underset{\sim}{x})$ given by $\ell(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n})'$. The gradient represents the direction cosines of the normal to the tangent hyperplane at point $\underset{\sim}{x}$ of the surface $f(\underset{\sim}{x})$. The method utilizes <u>steepest ascent</u> for a maximum and steepest

descent for a minimum, to increase the speed of approach to the optimum. Consider the matric

$$d^2 = (\underset{\sim}{x}-\underset{\sim}{y})' B(\underset{\sim}{x}-\underset{\sim}{y})$$

where B is a given matrix and $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are any two vectors. Then the <u>direction of steepest ascent</u> is the direction from the point x_0 to the ellipsoid

$$(x-x_0)' B(x-x_0) = k^2.$$

The following theorem gives an explicit form for optimization.

<u>Theorem 4.1</u>. For a function $f(\underset{\sim}{x})$, the maximum occurs in the direction $\delta(\underset{\sim}{x}_0)$ given by $\delta(\underset{\sim}{x}_0) = B^{-1}(\ell(\underset{\sim}{x}_0))$ where $\ell(\underset{\sim}{x}_0)$ is the gradient of $f(\underset{\sim}{x})$ at $\underset{\sim}{x}_0$. For proof and other relevant material the reader is referred to Crockett and Chernoff (1955).

5. Optimal Search Procedures

In optimum seeking methods, the aim is to design the most economic or shortest time consuming procedure. Suppose a function is to be explored over the points $x_1, x_2, \ldots, x_n$. Let $0 \leq x_i \leq 1$. Consider the following two situations with $n = 3$ in Figures 1 and 2, where the values of the function are given by verticle lines.

Figure 1

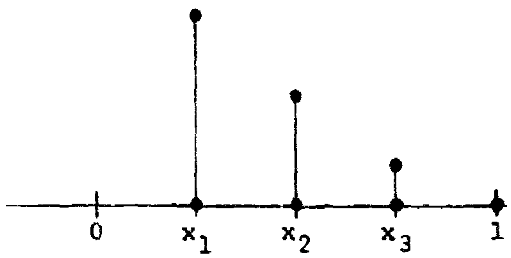
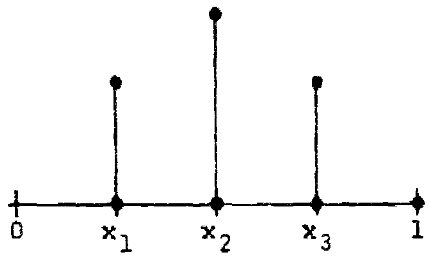


Figure 2



In Figure 1, the maximum may be in the interval $(0, x_2)$ and in Figure 2, it be in the interval (x_1, x_3). Such an interval is called the interval of uncertainty. In general the <u>interval of uncertainty</u> is (x_{k-1}, x_{k+1}). The length of the interval of uncertainty is given by $\ell(x_k, k)$. Several search plans based on the interval of uncertainty are given below.

<u>Minimax Search</u>. A plan which minimizes the maximum interval of uncertainty. That is,

$$\min_{x_1, x_2, \ldots, x_n} \max_{1 \leq k \leq n} \ell_n(x_k, k).$$

<u>Uniform Pairs Search</u>. It requires that the intervals chosen should be of uniform length. One such plan is to take

$$x_k = \frac{(1+\varepsilon)[\frac{k+1}{2}]}{\frac{n}{2} + 1} - \{[\tfrac{k+1}{2}] - [\tfrac{k}{2}]\}\varepsilon$$

where [a] denotes the integral part of a.

Other plans including the Fibonacci Search or Sequential Search plan, which is based on the Fibonacci sequence, and Golden Section Search plan, are also used in practice. For literature on optimum seeking methods, see Wilde (1964). An important class of optimum seeking procedures is concerned with optimizing the regression function in statistics. Such procedures have become known as Response Surface Methods. We shall discuss some elements of this methodology in the next section.

6. Response Surface Methods

The response surface methodology was developed to solve some problems in chemical investigation. However, its use became universal and in simulation methodology response surface techniques are very commonly used. The problem can be stated as follows. Let a region R, of k dimensions be called the factor space of with points $\underset{\sim}{x} = (x_1, x_2, \ldots, x_k)'$. Let the mean, μ of a response y_u depend on the factors $\underset{\sim}{x}_u$ through the function ϕ.

$$\mu_u = \phi(\underset{\sim}{x}_u). \tag{6.1}$$

Let y_u have variance σ^2. The problem then is to find a point $\underset{\sim}{x}^o$ in the smallest number of experiments so as to

optimize μ_u over the region R^k.

This classical problem was stated by Hotelling (1941) and Friedman and Savage (1947). Box and Wilson (1951) provided the basic framework to develop optimal response surface designs and their techniques have found considerable use in many applications. Myers (1971) has collected the available material in a book on response surfaces. We discuss elements of response surface methodology based on the paper of Box and Wilson. One of their major contributions was to develop new types of designs in place of complete factorial designs.

Let the distance, r, from the origin to the point $\underset{\sim}{x}$ be Euclidean, with

$$r^2 = \sum x_i^2. \tag{6.2}$$

The object here is to choose $\underset{\sim}{x}$ in such a way that

$$\phi(\underset{\sim}{x}) - \phi(0) \tag{6.3}$$

is maximized with the constraints in (6.2).

Using Lagrange's method, we maximize

$$\psi(\underset{\sim}{x}) = \phi(\underset{\sim}{x}) - \phi(0) - \frac{1}{2} \lambda \sum x_i^2 . \tag{6.4}$$

The stationary solution is given by equating to zero the partial derivatives with respect to x_i. We have

$$\lambda x_i = \frac{\partial \phi}{\partial x_i} (\underset{\sim}{x}). \tag{6.5}$$

Squaring and summing over all i and simplifying, we get

$$\lambda = \frac{1}{r} \left\{ \sum_{i=1}^{n} \left(\frac{\phi(\underset{\sim}{x})}{\partial x_i} \right)^2 \right\}^{1/2} \tag{6.6}$$

That is, the maximizing point should have coordinates proportional to the derivatives of ϕ.

Suppose the conditions of Taylor's expansion for $\phi(\underset{\sim}{x})$ in the neighborhood of the origin hold, then $\phi(\underset{\sim}{x})$ can be expanded to linear, quadratic and higher order terms. If we assume that second and higher order terms in the expansion of ϕ are zero, then, $\phi(\underset{\sim}{x})$ is approximated by a linear function of the following type:

$$\phi(\underset{\sim}{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{6.7}$$

Then,

$$\frac{\partial \phi(\underset{\sim}{x})}{\partial x_i} = \beta_i, \qquad i = 1, 2, \ldots, k \tag{6.8}$$

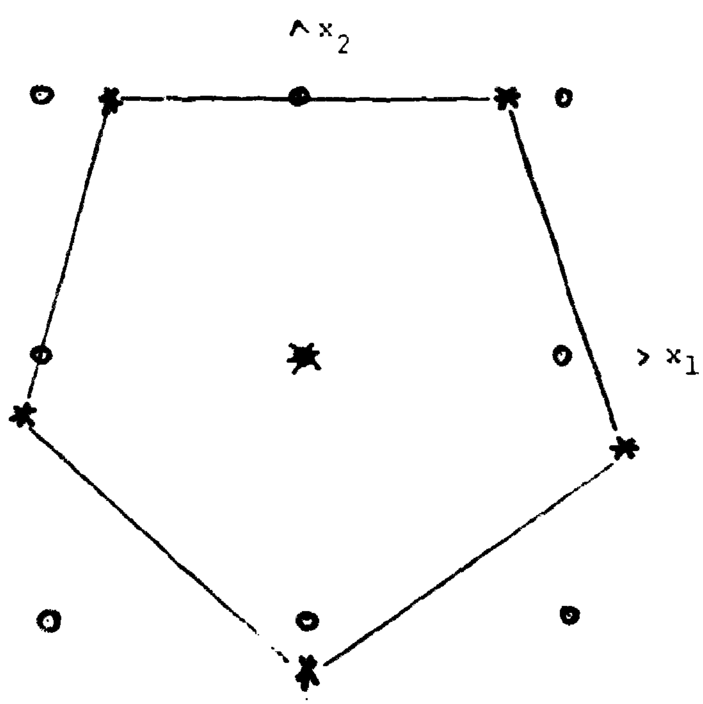
and the optimal x_i are proportional to β_i. Similarly expressions involving coefficients of linear and quadratic terms can be obtained if the Taylor's expansion of $\phi(\underset{\sim}{x})$ does not contain third and higher order terms. The move along the derivatives of the response function gives the steepest ascent approach to a maximum.

For the sake of clarity of presentation, we assume $k = 2$. Suppose $\phi(\underset{\sim}{x})$ has third and higher order derivatives zero. Hence we represent $\phi(\underset{\sim}{x})$ as follows:

$$\phi(\underset{\sim}{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 \tag{6.9}$$

Using the usual least squares theory, the regression equations (6.9) can be estimated by at least six or more points, since there are six unknown constants. As a rule, one would consider a complete 3x3 factorial experiment

Figure 3



with nine points so as to provide estimates for the quadratic regression (6.9). However, Box and Wilson provided a design, not of the factorial type which has five points on the vertices of a pentagon and the sixth at the origin. Such a design would give the estimates of the coefficients in the regression model and hence about the derivatives. These estimates then can be used to define the path of steepest ascent.

Several designs such as fractional factorials have also been used in response surface techniques and are available in text books on design of exepriments, for example, see Kempthorne (1978) and Myers (1971) providing a large number of new designs which are commonly applied in response surface methodology.

7. Optimal Design of Regression Experiments

The theory of optimal design of regression experiments is concerned with choosing the levels of the independent variable x for the model

$$y = f(x)$$

so as to optimize a certain function of parameters to be estimated in the model. We have given several optimality criteria as commonly used in optimal design theory in

Section 2. In simulation studies such criteria assume further importance since the design of a simulation may require several replications in a given problem. There is an extensive literature on optimality of designs. For a recent survey, see Federov (1972). Reviews of various other aspects of optimal designs have been presented more recently in the statistical literature. A review of D-optimality for regression designs has been given by St. John and Draper (1975) with an extensive bibliography.

A typical problem of optimal design theory is of the following type.

Example:

Consider the simple linear regression model

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i,\ i = 1, 2, \ldots, n \tag{7.1}$$

We assume that the errors ε_i are uncorrelated and have common variance σ^2. Let

$$\underset{\sim}{Y} = (y_1, y_2, \ldots, y_n)',\ \theta = (\theta_0, \theta_1)' \tag{7.2}$$

and

$$\underset{\sim}{x} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}. \tag{7.3}$$

Using the general linear model and results in Section 2, we find

$$\underset{\sim}{S} = \underset{\sim}{X}'\underset{\sim}{X} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \tag{7.4}$$

and

$$S^{-1} = a \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \tag{7.5}$$

$$\underset{\sim}{X}'\underset{\sim}{Y} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} \tag{7.6}$$

where

$$a = \frac{1}{n\sum(x_i-\bar{x})^2} \tag{7.7}$$

The estimates are given by

$$\hat{\underset{\sim}{\theta}} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix} = a \begin{pmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ -\sum x_i \sum y_i + n\sum x_i y_i \end{pmatrix} \tag{7.8}$$

and

$$\mathrm{Cov}(\hat{\theta}_0, \hat{\theta}_1) = -a\sigma^2 \sum x_i \tag{7.9}$$

$$V(\hat{\theta}_0) = a\sigma^2 \sum x_i^2 \tag{7.10}$$

$$V(\hat{\theta}_1) = a\sigma^2 n^{-1} \tag{7.11}$$

Suppose $V(\hat{\theta}_1)$ is to be minimized to obtain optimal x_i's. That is, the optimization problem is to maximize

$$\sum(x_i-\bar{x})^2. \tag{7.12}$$

Assuming that x's are between -1 and 1, the solution to the above problem is that x's should be placed at -1 and 1, half at each place to make (7.12) a maximum. For D-optimality, we maximize the determinant of S. That is, again we maximize

$$n\sum(x_i-\bar{x})^2. \tag{7.13}$$

Hence the same answer obtains as in minimizing the variance of $\hat{\theta}_1$.

Comparisons of Optimality Criteria

G-optimality (minimax optimality) was introduced by Smith (1918) and was exploited by Kiefer and Wolfowitz (1959). Wald (1943) used the criterion of D-optimality - in some other context and was so named by Kiefer and Wolfowitz (1959). One of the most important results in optimal design theory is the equivalence and characterizations of G-optimality and D-optimality under various conditions. This was established by Kiefer and Wolfowitz. Recently such results have also been extended to non-linear models by White (1973). Various computer algorithms

to generate D-optimum designs are available in the literature. Essentially the algorithm of Federov (1972), requires the following steps:

1) Select any non-degenerate starting design,
2) Compute the dispersion matrix,
3) Find the point of maximum variance,
4) Add the point of maximum variance to the design, with measure proportional to its variance
5) Update the design measure.

For further details, the reader is referred to the exposition by St. John and Draper (1975).

8. References

A. Classical Methods

Bloomfield, Peter and Watson, Geoffrey S. (1975). The inefficiency of least squares. Biometrika 62, 121-128.

Chanda, K. C. (1962). On bounds of serial correlations. Ann. Math. Statist. 33: 1457-1460.

Chattobadhyay, A. K., Pillai, K. C. S. and Li, H. C. (1976). Maximization of an integral of a matrix function and asymptotic exhausions of distributions of latent roots of two matrices. Ann. Statist. 4, 476-806.

Chaundy, Theodore (1935). The Differential Calculus. Oxford University Press.

Chernoff, Herman (1973). Some measures of discrimination between normal multivariate distributions with unequal covariance matrices. Proceedings of Multivariate Analysis-III. Editor, P. R. Krishnaiah, Academic Press, Inc., New York, 334-337.

Courant, R. (1936). Differential and Integral Calculus. Interscience, New York.

Dixon, L. C. W. (1972). Optimization in Action. Academic Press, New York.

Hancock, Harris (1917). Theory of Maxima and Minima. (Dover, New York, 1960).

Ireland, C. T. and Kullback, S. (1968). Contingency tables with given marginals. Biometrika 55, 179-188.

Lagrange, Conte Joseph Louise (1760-61). Essaid'une nouvella methode pour determiner les maxima et less minimis. Miscellanea Taurinensiz 1: 356-357, 360.

Lagrange, Conte Joseph Louise (1759). Researches sur la methode de maximis et minimis. Miscellanea Taurinensiz 1: 3-20.

Okamoto, M. and Kanazawa, M. (1958). Minimization of eigenvalues of a matrix and optimality of principal components. Ann. Math. Statist. 39: 859-863.

Walsh, G. R. (1975). Methods of Optimization. Wiley & Sons, New York.

B. Numerical Procedures

Beltrami, E. J. (1970). An Algorithmic Approach to Non-linear Analysis and Optimization. Academic Press, New York.

Box, M. J. (1966). A comparison of several current optimization methods and the use of transformations in constrained problems. The Computer Journal 2, 66-77.

Businger, P. and Golub, G. H. (1965). Linear least squares solution by Householder transformation. Numerishe Mathematik. 7: 269-276.

Carney, T. M. and Goldwyn, R. M. (1967). Numerical experiments with various optimal estimators. Journal of Optimization Theory and Applications. 1: 113-130.

Cauchy, A. (1847). Methode' generale pour la resolution des system d'equations simultan. C. R. Acad. Sci. Paris. 25: 536-538.

Crockett, J. B. and Charnoff, H. (1955). Gradient methods of maximization. Pacific J. Math. 5: 33-50.

Curry, H. B. (1944). The method of steepest descent for non-linear minimization problems. Quart. App. Math. 2: 258-261.

Dyer, P. and McReynolds, S. (1969). Extension of square root filtering to include process noise. Journal of Optimizing Theory and Applications, 3: 444-458.

Fitzgerald, R. J. (1966). A gradient method for optimizing stochastic systems. Recent Advances in Optimization Techniques. (A. Lavi and T. Vogel, Editors), John Wiley & Sons, New York.

Forsythe, G. (1955). Computing constrained minima with Lagrange multipliers. J. Soc. Indust. App. Math. 3: 173-178.

Hartley, H. O. (1961). The modified Gauss-Newton method for the fitting of non-linear regression function. Technometrics 3: 269-280.

Householder, A. S. (1953). Principles of Numerical Analysis. McGraw-Hill Book Co., New York.

Kale, B. K. (1962). On the solutions of likelihood equation by iterative processes - the multiparameter case. Biometrika 49, 479-486.

Kunz, H. P., Tzschach, H. G. and Zehnder, C. A. (1968). Numerical Methods of Mathematical Optimization. Academic Press.

Mortensen, R. E. (1968). Maximum likelihood recursive non-linear filtering. Journal of Optimization Theory and Applications 2: 386-394.

Olsson, D. M. and Nelson, Lloyd S. (1975). The Nelder-Mead simplex procedure for function minimization. Technometrics 17: 45-52.

Ostrowoski, A. M. (1960). Solutions of equations and systems of equations. Academic Pres, New York.

Ostrowoski, A. M., Contribution to the theory of the method method of steepest descent. Arch. Rational Mach. Anal. 26 (1967) 257-280.

Powell, M. J. D. (1964). An efficient method of finding the minimum of a cuntion of several variables without calculating derivatives. The Computer Journal, 7, 155-162.

Ralston, A. (1965). A First Course in Numerican Analysis. McGraw Hill Book Co., New York.

Shanno, D. R. (1970). An accelerated gradient projection method for linearly constrained non-linear estimation. J. Soc. Ind. App. Math. 13, 322-334.

Stuart, A. (1958). Iterative solutions of likelihood equations. Biometrics 14, 128-130.

Traub, J. F. (1964). Iterative Methods for Solution of Equations. Prentice-Hall, Inc., New York.

Wolfe, P. (1959). The secant method for simultaneous non-linear equations. Comm. ACM. 2: 12-13.

C. Mathematical Programming

Alldredge, J. R., and Armstrong, D. W. (1974). Maximum likelihood estimation for the multinomial distribution using geometric programming. Technometrics 16, 585.

Arthanari, T. S., and Dodge, Yadolah (1981). Mathematical Programming in Statistics. J. Wiley & Sons, New York.

Boot, John C. G. (1964). Quadratic Programming. North-Holland Publishing Co., Amsterdam.

Box, M. J., Davies, D. and Swann, W. H. (1969). Non-linear Optimization Techniques. Oliver and Boyd, Edinburgh.

Dantzig, G. B. (1963). Linear Programming and Extensions. Princeton University Press, Princeton, N. J.

Francis, Richard L. (1971). On relationships between the Neyman-Pearson problem and linear programming. Optimizing Methods in Statistics. (Editor, J. S. Rustagi) Academic Press, New York. 259-279.

Hadley, G. (1962). Linear Programming. Addison-Wesley, Reading, Massachusetts.

Hadley, G. (1964). Non-linear and Dynamic Programming. Addison-Wesley, Reading, Massachusetts.

Hartley, H. O. (1961). Nonlinear programming by the simplex method. Econometrics, 29: 223-237.

Kiountouzis, E. A. (1973). Linear programming techniques in regression analysis. App. Statist. 22, 69.

Mangasarian, O. L. (1969). Nonlinear Programming. McGraw Hill Co., New York.

Nelder, J. A. and Plead, R. (1964). A simplex method for function minimization. Computer Journal 7: 308-313.

Rao, M. R. (1971). Cluster analysis and mathematical programming. J. Am. Statist. Assoc. 66: 622.

Rustagi, J. S., Editor (1979). Optimizing Methods in Statistics. Academic Press, New York.

Rustagi, J. S., Editor(1971). Optimizing Methods in Statistics. Academic Press, New York.

Rustagi, J. S. (1968). Dynamic programming model of patient care. Math. Biosc. 141-149.

Spendley, W. Hext and Hunsworth, F. R. (1962). Sequential application of simplex designs in optimization and EVOP. Technometrics 4: 441-461.

Wagner, D. H. (1969). Nonlinear functional versions of the Neyman-Pearson lemma. SIAM Rev. 11, 52.

Wagner, H. M. (1959). Linear programming and regression analysis. Journal of American Statist. Assoc. 54: 206-212.

Wagner, H. M. (1962). Nonlinear regression with minimal assumptions. Journal of American Statist. Assoc. 57: 572-578.

Whittle, Peter (1971). Optimization under Constraints. Wiley-Interscience, New York.

D. Stochastic Approximation Methods

Albert, A. E. and Gardner, Jr., L. A. (1967). Stochastic Approximation and Non-linear Regression. M.I.T. Press, Cambridge, Mass.

Blum, J. R. (1958). A note on stochastic approximation. Proc. Amer. Math. Soc. 9: 404-407.

Blum, J. R. (1954). Approximation methods which converge with probability one. Ann. Math. Statist. 25: 382-386.

Blum, J. R. (1954). Multidimensional stochastic approximation methods. Ann. Math. Statist. 25: 737-744.

Burkholder, D. (1956). On a class of stochastic approximation processes. Ann. Math. Statist. 27: 1044-1059.

Chung, Kai-Lai (1954). On a stochastic approximation method. Ann. Math. Statist. 25: 463-483.

Derman, C. and Sacks, J. (1959). On Dvoretzky's stochastic approximation theorem. Ann. Math. Statist. 30: 601-606.

Dupac, V. (1965). A dynamic stochastic approximation method. Ann. Math. Statist. 36: 1695-1702.

Dvoretzky, A. (1956). On stochastic approximation. Proceedings Third Berkeley Symp. Math. Statist. and Prob. 1: 39-56.

Fabian, V. (1971). Stochastic approximation. Optimizing Methods in Statistics. (Ed., J. S. Rustagi), Academic Press, New York, 439-470.

Fabian, V. (1960). Stochastic approximation methods. Czechoslovak Math. 10: 123-159.

Gladyschev, E. G. (1965). On stochastic approximation. Theor. Probability Appl. 10: 275-278. (English Translation).

Hodges, J. and Lehmann, E. (1956). Two approximations to the Robbins-Monro process. Proc. Third Berkeley Symp. Math. Statist. and Prob. 1: 95-104.

Kallianpur, G. (1954). A note on the Robbins-Monro stochastic approximation method. Ann. Math. Statist. 25: 386-388.

Kiefer, J. and Wolfowitz, Jr. (1952). Stochastic estimation of the maximum of a regression function. Ann. Math. Statist. 23: 462-466.

Krasulina, T. P. (1962). A note on some stochastic approximation processes. Theor. Probability Appl. 7: 108-113. (English Translation).

Robbins, H. and Monro, S. (1951). A stochastic approximation method. Ann. Math. Statist. 22: 400-407.

Sakrison, David J. (1964). A continuous Kiefer-Wolfowitz procedure for random processes. Ann. Math. Statist. 35: 590-599.

Venter, J. H. (1967). On convergence of the Kiefer-Wolfowitz approximation procedure. Ann. Math. Statist. 38: 1031-1036.

Wasan, M. T. (1969). Stochastic Approximation. Cambridge University Press, London.

Wolfowitz, J. (1956). On stochastic approximation methods. Ann. Math. Statist. 27: 1151-1156.

Wolfowitz, J. (1952). On the stochastic approximation method of Robbins and Monro. Ann. Math. Statist. 23: 457-461.

E. Optimum Seeking Methods and Response Surface Methodology

Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum considerations. J. Roy. Statist. Soc. Series B 13: 1-38.

Birnbaum, David and Doray, Rolan (1978). Optimal allocation of resources using multidimensional search techniques. Symposium on Modelling and Simulation Methodology. Weizmann Institute of Science, Israel.

Hooke, R. and Jeeves, T. A. (1961). Direct search solutions of numerical and statistical problems. J. Assoc. Comput. Mach. 8, 212-219.

Karson, M. J. (1970). Design criterion for minimum bias estimation of response surfaces. J. Am. Statist. Assoc. 65, 1565-1572.

Kiefer, J. (1953). Sequential minimax search for a maximum. Proc. Amer. Math. Soc. 4, 502-506.

Myers, R. H. (1971). Response Surface Methodology. Allyn and Bacon, Inc., New York.

Schewfel, H. P. (1978). Direct search for optimal parameters within simulation models. Symposium on Modelling and Simulation Methodology, Weizmann Institute of Science, Israel.

Shah, B. V., Buehler, R. J. and Kempthorne, O. (1964). Some algorithms for minimizing a function of several variables. J. Soc. Indust. App. Math. 12, 74-92.

Wilde, Douglas J. (1964). Optimum Seeking Methods. Prentice-Hall, New York.

F. Optimal Design Theory

Evans, James W. (1979). Computer augmentation of experimental designs to maximize $|X'X|$, Technometrics, 21: 321-330.

Federov, V. V. (1977). Theory of Optimal Experiments. Academic Press, London.

Kempthorne, O. (1975). Design and Analysis of Experiments. Krieger Publishing Company, New York.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. Ann. Math. Statist. 30, 271-294.

Rao, J. N. K. (1979). Optimization in the design of sample surveys. Optimizing Methods in Statistics. (Editor, J. S. Rustagi). Academic Press, New York, 419-434.

Rustagi, J. S. (1976). Variational Methods in Statistics. Academic Press, New York.

St. John, R. C. and Draper, N. R. (1975). D-optimality for regression designs: A survey. Technometrics 17, 15-23.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. Biometrika 12, 1-85.

Srivastava, J. N. (Editor) (1975). A Survey of Statistical Design and Linear Models. North-Holland Publishing Company, Amsterdam.

Wald, A. (1943). On the efficient design of statistical investigations. Ann. Math. Statist. 14, 134-140.

G. Miscellaneous

Brown, G. H. (1979). An optimization criterion for linear inverse estimates. Technometrics 21: 575-579.

Friedman, M. and Savage, A. J. (1947). Selected Techniques of Statistical Analysis. McGraw-Hill, New York.

Hotelling, H. (1941). Experimental determination of the maximum of an empirical function. Ann. Math. Statist. 12, 20-45.

Kleijnen, J. P. C. (1974). Statistical Techniques in Simulation. Marcel Dekker, Inc., New York.

Milstein, Jaime (1979). Modelling and parameter identification of insulin action on gluconeogenesis: An integrated application of advanced modelling methods. Methodology in Systems Modelling and Simulation. (Ziegler, B. P. et al. Editors). North-Holland Publishing Company, Amsterdam, 271-288.

Rustagi, J. S. and Zanakis, S. (Editors) (1981). Optimization in Statistics. North-Holland Publishing Company, Amsterdam. (To appear).

Rustagi, J. S. (Editor) (1978). Special Issue on **Optimization in Statistics**. Commun. Statist. - Simula.Computa. B7(4).

Rustagi, J. S. (1957). Minimizing and maximizing certain integral with statistical applications. Ann. Math. Statist. 28, 309-328.

Tapia, Richard A. and Thompson, J. R. (1978). Nonparametric Probability Density Estimation. Johns Hopkins University Press, Baltimore, Maryland.

Vogel, Joseph S. (1978). Systems simulation and optimization in the presence of nonlinear constraints. Symposium on Modelling and Simulation Methodology. Weizmann Institute of Science, Israel.

Weeks, J. K. and Fryer, J. S. (1977). A methodology for assigning minimum cost due-dates. Management Sciences 23: 872-881.

Zeigler, Bernard P., et al. (Editors) (1979). Methodology in Systems Modelling and Simulation. North-Holland Publishing Company, Amsterdam.

Zeigler, B. P. (1976). Theory of Modelling and Simulation. John Wiley & Sons, Inc. New York.